

<https://helda.helsinki.fi>

---

## Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity

Oversti, Sanni

2017-07-21

---

Oversti , S , Onkamo , P , Stoljarova , M , Budowle , B , Sajantila , A & Palo , J U 2017 , ' Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity ' , Scientific Reports , vol. 7 , 6193 . <https://doi.org/10.1038/s41598-017-05673-7>

---

<http://hdl.handle.net/10138/215230>

<https://doi.org/10.1038/s41598-017-05673-7>

---

cc\_by

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# SCIENTIFIC REPORTS

OPEN

## Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity

Sanni Översti<sup>1</sup>, Päivi Onkamo<sup>1</sup>, Monika Stoljarova<sup>2</sup>, Bruce Budowle<sup>3,4</sup>, Antti Sajantila<sup>5</sup> & Jukka U. Palo<sup>5,6</sup>

In Europe, modern mitochondrial diversity is relatively homogeneous and suggests an ubiquitous rapid population growth since the Neolithic revolution. Similar patterns also have been observed in mitochondrial control region data in Finland, which contrasts with the distinctive autosomal and Y-chromosomal diversity among Finns. A different picture emerges from the 843 whole mitochondrial genomes from modern Finns analyzed here. Up to one third of the subhaplogroups can be considered as Finn-characteristic, i.e. rather common in Finland but virtually absent or rare elsewhere in Europe. Bayesian phylogenetic analyses suggest that most of these attributed Finnish lineages date back to around 3,000–5,000 years, coinciding with the arrival of Corded Ware culture and agriculture into Finland. Bayesian estimation of past effective population sizes reveals two differing demographic histories: 1) the 'local' Finnish mtDNA haplotypes yielding small and dwindling size estimates for most of the past; and 2) the 'immigrant' haplotypes showing growth typical of most European populations. The results based on the local diversity are more in line with that known about Finns from other studies, e.g., Y-chromosome analyses and archaeology findings. The mitochondrial gene pool thus may contain signals of local population history that cannot be readily deduced from the total diversity.

The current genetic diversity of any population is a result from the actions of four evolutionary forces (drift, mutation, migration and selection). Reconstructing the past from modern genetic data can thus be challenging. The confounding signals resulting from these different forces can be minimized to some extent, e.g., by marker choice (ignoring selection) or by assuming steady mutation rates, but genetic drift and gene flow are more difficult to address. This issue is especially problematic in human populations due to relatively high migration rates and very short genetic distances between populations with evolutionarily young species age and small founding populations. Yet, the two latter forces are the most important ones when inferring population history as they reflect changes in population size and migration patterns during different phases of history.

An increasingly popular approach to elucidate different demographic events is to analyze past diversity directly from archaeological samples, i.e., ancient DNA analyses. Despite recent technical and analytical advances this approach is still relatively expensive, time-consuming, error-prone and often involves very limited size data sets. An alternate approach is to use modern genetic data and indirect means to deduce the past. These studies could include simulation methods that allow evaluation of the likelihoods of alternative demographical scenarios or *a priori* selection of sampled individuals based on, e.g., locality or surnames (for review see ref. 1). In the study

<sup>1</sup>Department of Biosciences, University of Helsinki, PO Box 56, FI-00014, Helsinki, Finland. <sup>2</sup>Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia. <sup>3</sup>Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, Texas, USA. <sup>4</sup>Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia. <sup>5</sup>Department of Forensic Medicine, University of Helsinki, PO Box 40, FI-00014, Helsinki, Finland. <sup>6</sup>Forensic Genetics Unit, National Institute for Health and Welfare, PO Box 30, FI-00271, Helsinki, Finland. Correspondence and requests for materials should be addressed to S.Ö. (email: [sanni.oversti@helsinki.fi](mailto:sanni.oversti@helsinki.fi))

N	Reference	GenBank ID numbers	Origin
93	24	n/a	n/a
192	19	AY339402–AY339593	Northern Ostrobothnia, central Ostrobothnia, Kainuu, and northern Savo
94	23	Supplementary Table S1	n/a
293	22	Supplementary Table S1	Tampere
64	21	JX171077–JX171140	Northern Finland
79	20	KC763372–KC763450	Northern Ostrobothnia
28	*, **43, 68	AY195779*, JF837819**, EU753433, EU784076, FJ543390, FJ801039, GQ176284, GU206811, GU391321, GU949563, HM116534, HM856585, HQ022823, HQ658464, HQ840516, JF298814, JF813785, JF813786, JF903928, JQ086344, KF466256, KF631316, KM213522, KP782041, KR779775, KR712271, KR919601, KR902539	n/a
<b>TOTAL 843</b>			

**Table 1.** Modern human complete mtDNA genomes from Finland (N = 843). Sequences were obtained through GenBank searches, 1000 Genomes data and from previously published articles.

herein the rationale followed was exploring demographic signals in two different spatiotemporal layers of mitochondrial diversity by segregating the modern data into “local” and “immigrant” components<sup>2</sup>. This approach revealed demographic history patterns that remained hidden within the total population variation.

The population of Finland was explored because of its rather well-studied, and peculiar history. This history is branded by distinct colonization waves, high drift due to small size and isolation-by-distance<sup>3,4</sup>. Y-chromosomal and genome wide studies support that Finnish genetic diversity patterns differ compared with most other West Eurasian populations (see e.g. ref. 5). The Y chromosomal diversity is especially reduced and distinctive, showing a clear east-west divergence over short geographic distances within the country<sup>4–8</sup>. This genetic difference may have arisen from an ancient border between two different immigration waves that arrived at different times and could represent two different modes of subsistence, i.e., hunter-gathering (east) and animal husbandry/farming (west)<sup>9</sup>. A similar but less dramatic border has been observed with the autosomal data<sup>8</sup> and even in genes with pharmacogenetic relevance<sup>10,11</sup>.

Whereas the reduced and distinctive Y-chromosomal diversity suggests a history of population bottlenecks and immigration waves, most mtDNA studies have observed an overall pattern similar to that of most other European populations, i.e. high diversity and low geographical segregation of diversity<sup>4,6,9,12–14</sup>. Finns carry mostly European haplogroups H, I, J, K, T, V, W and X, but also to some extent Asian haplogroups D and Z<sup>12</sup>. However, some studies have observed weak geographical patterns in the Finnish mtDNA diversity<sup>4,9,13</sup>. The Finns show vestiges of an ancient border between the Mesolithic hunter-gatherer (U and V, see Discussion) and Neolithic farmer mtDNA haplogroups (H, J, K and T), a pattern previously identified in their Y-chromosomes<sup>9</sup>.

The demographic signals in the two notable components of mtDNA variation, each with different histories in Finland, were studied. Based on the geographical distribution, mtDNA lineages were searched in public repositories. Unlike earlier studies focusing on Finland, this study aimed to achieve a fine resolution picture by analyzing only full mtDNA genome sequence data. A significant proportion of subhaplogroups in a Finnish data set, up to one third, are rare or absent outside Finland, a mtDNA component henceforth called “Finn-characteristic” lineages. The results indicate that separate analyses of local and immigrating lineages allowed detection of different population history signals. In this study, this signal was most evident in pre-historic effective population sizes. Most importantly, these patterns have not been detected within total population variation studies.

## Material and Methods

**Database searches and haplogroup assignment.** Complete modern mtDNA genomes from Finland were obtained through GenBank searches, HmtDB searches<sup>15,16</sup> and from published literature, including information on the population of origin. Altogether 843 complete Finnish sequences were retrieved (Table 1 and Supplementary Dataset File). Sequences were aligned with Muscle v3.8.31<sup>17</sup> to the revised Cambridge reference sequence, rCRS<sup>18</sup>. Finnish samples used in this study might be slightly biased towards the northern parts of Finland: dataset contained samples from northern Ostrobothnia, central Ostrobothnia, Kainuu and northern Savo in refs 19 and 20 and Northern Finland in ref. 21. Nevertheless, the samples from Tampere<sup>22</sup> can be considered more representative of southern and western parts of the country. More detailed place of origin was not available for all samples<sup>23,24</sup>.

It is worth noting that altogether 335 of the Finnish sequences (published in refs 19–21) were obtained by sequencing mtDNA genomes initially showing variation by conformation-sensitive gel electrophoresis (CSGE). The potential influence of this approach on the results is discussed below (see Results).

The mtDNA haplogroup information for all sequences was inferred from the sequence data both by using Haplofind<sup>25</sup> and HaploGrep2<sup>26</sup>, in order to gain greater certainty for the haplogroup definitions. Haplofind is based on PhyloTree version 16<sup>27</sup> HaploGrep2 is based on PhyloTree version 17<sup>27</sup>. For less clear haplogroup assignments the results from HaploGrep2 were considered to be more reliable and relied upon in this study.

To identify mitochondrial lineages characteristic for Finns BLAST (Basic Local Alignment Tool<sup>28</sup>) searches were conducted against sequences deposited in GenBank to determine the geographical distribution of haplotypes in the different (sub)haplogroups. The GenBank database contains around 30,000 complete human

Scheme number	Composition	Best model according to the PartitionFinder	Best model according to the Bayesian analysis with BEAST v1.8.2	Substitution rate this study ( $\mu$ /Site/Year $\times 10^{-8}$ )	Substitution rate Rieux et al. 2014 ( $\mu$ /Site/Year $\times 10^{-8}$ )
1	Nucleotides 1–576 and 16,024–16,569	GTR+ $\Gamma$ + Inv	TrN+ $\Gamma$ + Inv	20.56 [12.49, 29.01]	31.43 [22.56, 40.31]*
2	rRNA, tRNA	HKY+ $\Gamma$ + Inv	HKY+ $\Gamma$ + Inv	1.58 [0.94, 2.30]	1.01 [0.76, 1.27]
3	PC1, PC2	TrN+ $\Gamma$ + Inv	TrN+ $\Gamma$ + Inv	1.80 [1.17, 2.40]	0.76 [0.57, 0.94]
4	PC3	TrN+ $\Gamma$ + Inv	TrN+ $\Gamma$ + Inv	3.04 [1.92, 4.21]	3.32 [2.57, 4.07]

**Table 2.** The best-fit nucleotide substitution model and substitution rate estimation for each partitioning schemes. GTR = General Time Reversible<sup>33</sup>, HKY = Hasegawa-Kishino-Yano<sup>69</sup>, TrN = Tamura-Nei<sup>34</sup>,  $\Gamma$  = Gamma distribution and Inv = Invariant sites. In all the models the number of gamma categories was set to 4. Substitution rates are the mean with 95% highest posterior density interval. For the comparison, the substitutions rates obtained from Rieux *et al.*<sup>38</sup> also are presented. \*Mutation rate estimated for hypervariable region 1 and 2 (HVR1 + HVR2).

mitochondrial genomes (estimated from GenBank <http://www.ncbi.nlm.nih.gov/genbank/> and Human Mitochondrial database <http://www.hmtdb.uniba.it/hmdb/> 15.03.2017). This resource thus is a larger database than that of the commonly used PhyloTree database (24,275 complete or nearly complete genomes; 15.3.2017). PhyloTree and GenBank databases overlap, with over 90% of complete genomes deposited in PhyloTree also housed in GenBank.

As a query sequence for each BLAST search one complete mitochondrial sequence from each possible Finnish characteristic haplogroup was used. For each search the maximum number of sequences to display was set to 500 and for all these 500 sequences the haplogroup was determined by HaploGrep2. Haplogroups, which contained less than four sequences in our Finnish sample (such as H1f1a, H1a8 and U5b2a1a1b) or which were known to be common in other populations (e.g. H2a1, U5b1b1a and V7a1), were excluded from the searches. The ethnic origins of the sequences were inferred from GenBank and/or from original publications. Estimated from the Human Mitochondrial DataBase, around 20% of complete mtDNA sequences in GenBank are of undefined origin (Supplementary Table S1). Note, however, that roughly one fifth of the GenBank sequences lack precise information on ethnic/geographical origin and the origin for those sequences is designated as “n/a” (see Results). Despite the uncertainty of the ethnic/geographical origin of these sequences, all sequences deposited in GenBank were included in the subsequent analysis to avoid an overestimation of the proportion of Finn-characteristic haplogroups. It is possible that these sequences also include Finnish samples, which in turn would indicate that the Finn-characteristic haplogroups could be even more restricted to Finns than it appears based on the results. The numbers of complete mtDNA sequences from nearby populations are presented in Supplementary Table S2.

Haplogroups, for which Finns constituted more than 75% of the sequences, were considered Finn-characteristic. In order to visualize the diversity within these haplogroups, Neighbor-Joining (NJ) trees were reconstructed with MEGA 6<sup>29</sup> using substitution model TrN+  $\Gamma$  with shape parameter  $\alpha = 0.02$ . The choice of the model was based on the best-fit model estimated with PartitionFinder<sup>30</sup> (TrN+  $\Gamma$ + Inv) and gamma shape parameter estimated by BEAST v1.8.2. Minor modifications of these values were necessary due to MEGA software limitations. Trees were rooted with L3a haplogroup sequence JN655813<sup>31</sup>. The branch-support values were obtained by bootstrapping (500 resamples). NJ trees are presented in the Supplementary Figs S1–S4.

As the 75% inclusion limit is arbitrary, it was necessary to assess whether the obtained results are heavily dependent on the cut-off limit set to define Finn-characteristic haplogroups. A more stringent cut-off limit of 90% also was applied to the data, and the results obtained with the different data sets were compared.

**Phylogenetic analyses.** In order to estimate the age of the haplogroups and past population sizes the Bayesian approach in BEAST v1.8.2<sup>32</sup> was used. For these analyses, the optimal partition scheme of the mtDNA variation and the best-fit nucleotide substitution model for each partition were estimated using the software PartitionFinder<sup>30</sup> assuming seven categories of nucleotides:

- 1) Nucleotides 1–576 (including hypervariable region 2)
- 2) Nucleotides 16,024–16,569 (including hypervariable region 1)
- 3) Protein coding positions at 1<sup>st</sup> codon (PC1)
- 4) Protein coding positions at 2<sup>nd</sup> codon (PC2)
- 5) Protein coding positions at 3<sup>rd</sup> codon (PC3)
- 6) Transfer RNAs (tRNAs)
- 7) Ribosomal RNAs (rRNAs)

PartitionFinder analyzes every scheme that includes these seven categories in all possible combinations. The best partitioning was obtained for four schemes (1 + 2, 3, 4 + 5 and 6 + 7). The most appropriate nucleotide substitution model for each scheme is presented in Table 2. The preferred model, the General Time Reversible model<sup>33</sup>, was too complex of a model for scheme 1 (non-coding control region nucleotide positions 1–576 and 16,024–16,569) with the subsequent Bayesian phylogenetic analyses. Therefore, the GTR model was changed into the more simplified Tamura-Nei model<sup>34</sup> and was used in all subsequent analyses (Table 2).

All the Bayesian phylogenetic analyses were performed in BEAST v1.8.2<sup>32</sup>. Each BEAST run was performed in the same way: single MCMC chain was ran for 40,000,000 steps, sampled every 4,000 steps and the first 10% was

discarded as a burn-in. For each of the following analyses four independent runs were performed, and these runs were combined with LogCombiner to achieve adequate effective sample size (ESS) values<sup>35</sup>. Appropriate effective sample size values (ESS > 200) for each parameter in the model were checked in Tracer v1.6<sup>36</sup>.

Substitution models and clock models were unlinked between the schemes in all analyses; but the tree topology was assumed to be the same between the four schemes. The most suitable demographic and clock models were determined by Bayes factors calculated from the marginal likelihoods<sup>37</sup>. The data were fitted into three demographic models (Bayesian skyline plot, exponential growth and constant population size models). The most appropriate demographic model based on Log<sub>10</sub> BF comparison was the constant population size model (Log<sub>10</sub> BF 1.92 and 2.22 compared to the BSP and exponential growth models, respectively). The constant population size model is considered to be most suitable for population level analyses (see <http://www.beast2.org/tree-priors-and-dating/> 15.03.2017). The most suitable molecular clock for the data was strict clock (Log<sub>10</sub> BF 4.05 over lognormal relaxed clock). The prior distributions for mutation rates for each of the four partitions were set based on Rieux *et al.* 2014<sup>38</sup>.

Because the tip calibration has been shown to yield more consistent results than node dating<sup>38</sup>, 12 previously published radiocarbon dated sequences were used as calibration points in our analyses (Supplementary Table S3). The tip date uncertainty was taken into account, as suggested by Rieux *et al.* 2014<sup>38</sup>. The 12 ancient sequences were from different time points as it has been shown that even a small sample of aDNA sequences with widely distributed and accurate ages are sufficiently powerful to produce precise estimates for mutation rate and timescales<sup>39</sup>.

**Bayesian Skyline Plots.** The effective population sizes for Finn-characteristic and non-Finn-characteristic haplogroups were compared by constructing Bayesian Skyline Plots (BSPs)<sup>40</sup> with BEAST v1.8.2<sup>32</sup>. The coding region (nucleotides 577–16,023) was used in BSP analyses. The best-fit substitution model was estimated with PartitionFinder<sup>30</sup>. The substitution model used was TrN+  $\Gamma$ + Inv with four gamma categories and partitioned into codon positions (1 + 2), 3. The tree prior used was piecewise-linear Skyline coalescent model with number of groups 10. The prior normal distribution for mutation rate was estimated based on six previously published mutation rates for mtDNA coding region<sup>41–46</sup>. Resulting mean mutation rate was  $\mu = 1.546 \times 10^{-8} \pm 3.675 \times 10^{-9}$  (SD) substitutions per base pair per year.

In order to test the robustness of analysis, BSPs were constructed for those Finn-characteristic haplogroups for which Finns constituted  $\geq 90\%$  of the sequences.

As an additional test for validity of obtained results BSP plots were constructed using HVR1 + 2 data (643 bp) from Neuvonen *et al.* 2015<sup>9</sup>, of which geographical origin information was available. In order to explore potential population structure effects, the BSPs were constructed separately for eastern Finland, western Finland and all data (see ref. 9) assuming the same run parameters as for the 1 + 2 partition scheme described above.

## Results

**Finnish mtDNA diversity.** A total of 843 Finnish mitochondrial genomes were retrieved from the databases. These data segregate into 240 subhaplogroups (Supplementary Table S4) with overall haplogroup frequencies similar with those published previously<sup>9,12,13</sup>. To evaluate a possible sampling bias, due to lack of precise information of the geographical origin of the Finnish samples used in this study, the main haplogroup frequencies were compared between this study and HVR1 + 2 data presented in Palo *et al.* 2009<sup>4</sup>. In Palo *et al.* 2009 the donor's reported place of residence was known, and the sampling covered Finland as a whole. There were no significant differences in the frequencies between these two studies, which supports that the data of this study can be considered a reasonable representative sample of Finland. (Supplementary Table S5).

Unlike most of the sequences, a significant portion of the Finnish sequences ( $n = 335$ ) deposited in GenBank were originally obtained by sequencing only representative mtDNA genomes that showed variation with CSGE. The reliability of these data therefore depends on the capability of the CSGE technique in detecting variation among genomes. This was not considered a significant source of bias. First, as with the sampling bias discussed above, congruence in haplogroup frequencies between this study and previous ones supports data coherence. Second, the CSGE method has shown to be able to resolve single base pair differences in K rkk  *et al.* 1998<sup>47</sup>.

Using the cut-off limit of 75% (Table 3), a total of 281 Finn-characteristic sequences belonging to 23 mitochondrial lineages was identified. These lineages constituted up to 33.3% of our whole data set and comprised both hunter-gatherer (haplogroups U and V) and farmer (haplogroups H, J and K) associated lineages<sup>48</sup>. With the more stringent cut-off limit of 90%, 20.8% of all mtDNAs were designated Finn-characteristic. The proportion of Finn-characteristic haplogroups within each dataset used<sup>19–24</sup> varied between 20.2–45.3% (Supplementary Table S6).

Based on the Bayesian phylogenetic analysis the mutation rates obtained in this study were similar with rates previously published by Rieux *et al.* 2014<sup>38</sup> (Table 2). Slight differences were observed: the mutation rate for PC1 + PC2 was slightly higher ( $\mu = 0.76 \times 10^{-8}$  mutations/site/year in Rieux *et al.* 2014 vs.  $\mu = 1.80 \times 10^{-8}$  mutations/site/year in this study) and the rate for nucleotides 1–576 and 16,024–16,569 (HVS1 + HVS2) was slightly lower compared to Rieux *et al.* 2014 rates. These differences are minor and probably arise from the different data sets used in these studies. For each of the schemes the base frequencies, transition/transversion ratios, proportion of the invariant sites and the shape parameter  $\alpha$  were estimated from the data (Supplementary Table S7). In the control region the transitional rate between pyrimidines has been shown to be higher than that between purines<sup>34</sup>. For protein-coding genes the mutation rate for guanine has been estimated to be very high compared to other parts of the mitochondrial genome<sup>49</sup> resulting in higher transitional rate between purines. The same phenomena were visible in these data. The gamma distribution shape parameter  $\alpha$  was the highest in the scheme 1 (nucleotides 1–576 and 16,024–16,569 with  $\alpha = 0.62 \pm 0.29$ ), and the proportion of invariants was the highest in scheme 4 (rRNAs and tRNAs, pInv =  $0.94 \pm 0.03$ ).



Haplogroup	N *	N Finn/% **	Others (GenBank ID, origin if available)	Prevalence in Finland based on this study (N <sub>Finns</sub> = 843)
H13a1a1d1	11	11/100.0	—	1.3
H1a2	26	21/80.8	EU687746.1 Saami Norway, HM775496.1 Sweden, EU130562.1 n/a, JQ702405.1 n/a, EST_27	2.5
H1f1	30	27/90.0	JQ704916.1 n/a, JQ705985.1 n/a, EST_15	3.2
H1n4	13	10/76.9	JQ702043.1 n/a, EST_45, EST_75	1.2
H3h1	17	14/82.4	KF162052 Denmark, KJ446376 = KF451393 Russia, EST_5	1.7
H5a1e	4	4/100.0	—	0.5
I1a1a1	5	5/100.0	—	0.6
I1a1a2	4	4/100.0	—	0.5
I2b	4	4/100.0	—	0.5
J1c2n1	8	8/100.0	—	0.9
K1c1c	22	18/81.8	KC170989 Kirov Russia, KF162649 Denmark, EU262720 n/a, EST_1	2.1
U5a2a1a	14	13/92.9	KF161330 Denmark	1.5
U5b1b1a1a	11	9/81.8	GU296598.1 Russia, JQ702500.1 n/a	1.1
U5b1b1a1a1	7	7/100.0	-	0.8
U5b1b2	30	23/76.7	JX026063.1 Sweden, JQ702152.1 Ireland, KF161244.1 Denmark, KF162959.1 Denmark, KF162429.1 Denmark, EU367993.1 n/a, EST_18	2.7
U5b1b2a	12	11/91.7	EST_107	1.4
V1a1a	6	6/100.0	—	0.7
V1a1a1	12	12/100.0	—	1.4
V5	10	9/90.0	JQ705942.1 n/a	0.7
V8	14	11/78.6	JQ705604.1 n/a, JQ703985.1 n/a, JQ702025.1 n/a	1.3
W1a	37	35/94.6	KF061034 England, EST_37	4.2
W1b	4	4/100.0	—	0.5
W1b1	14	14/100.0	—	1.7
Total	315	281/89.2	—	33.3

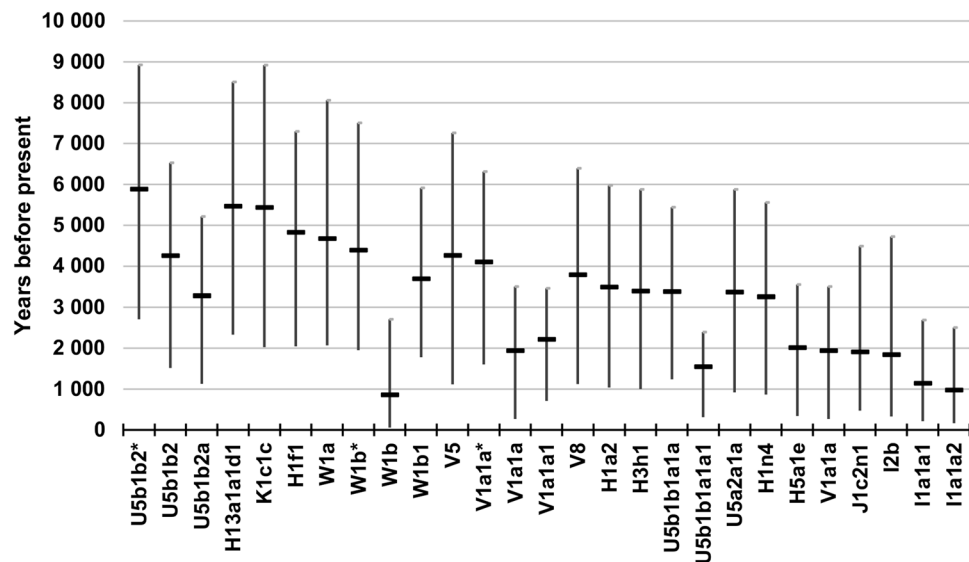
**Table 3.** Finn-characteristic haplogroups. Estonian samples (EST\_27, EST\_15, EST\_45, EST\_75, EST\_5, EST\_1, EST\_18, EST\_107 and EST\_37) from Stoljarova *et al.*<sup>70</sup>. n/a = ethnic/geographical origin for the sequence not available. \*Number of sequences belonging to the haplogroup. \*\*Number of Finnish sequences belonging to the haplogroup/the percentage of Finnish sequences in the haplogroup.

Ages for the Finn-characteristic haplogroups, estimated using the mutation rates above, are presented in Fig. 1 and Supplementary Table S8. The oldest subhaplogroup, U5b1b2\*, dates back almost 6,000 years (median 5,883 years with 95% highest posterior density [2,708; 8,923] years), while the youngest subhaplogroup, W1b, is estimated to be less than 900 years old (median 858 years with 95% highest posterior density [57; 2,710] years). For a notable proportion of subhaplogroups (i.e. haplogroups H1n4, U5b1b2a, U5a2a1a, U5b1b1a1a, H3h1, H1a2, W1b1, V8, V1a1a\*, U5b1b2, V5, W1b\*, W1a, H1f1, K1c1c and H13a1a1d1) the median age estimate ranged from 3,300–5,500 ybp.

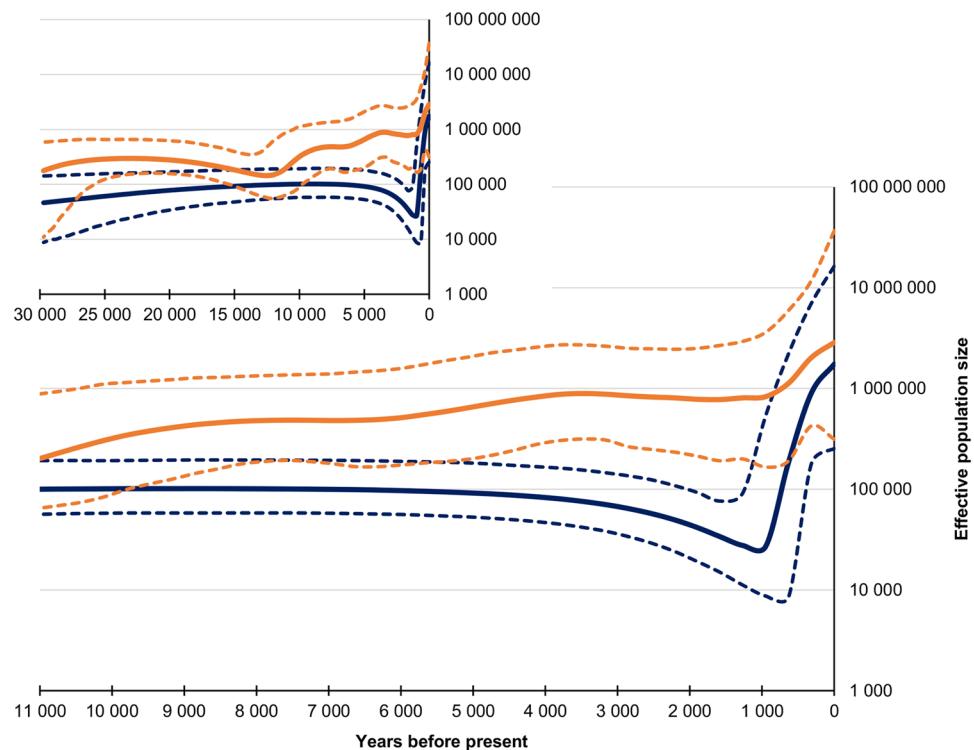
**Past population sizes: Bayesian skyline plots.** In order to analyze population size changes over time, BSPs were constructed for Finn-characteristic and for other haplogroups (Fig. 2). For the non-Finn-characteristic haplogroups the increase in effective population size ( $N_e$ ) was relatively stable until around 4,000 years before present when the increase stops. The  $N_e$  was constant between 4,000–1,000 years before present after which it started to increase rapidly. Instead, for the Finn-characteristic haplogroups, the  $N_e$  was relatively stable until it started to decrease around 5,000–4,000 years before present. The lowest point, below one third of the previous stable  $N_e$ , was reached around 1,000 ybp. The growth for Finn-characteristic haplogroups during the last 1,000 years was exponential, and the present  $N_e$  is over sixty times larger compared to the  $N_e$  of the lowest point. The present  $N_e$  for Finn-characteristic haplogroups is two thirds of the  $N_e$  for other haplogroups.

## Discussion

This study shows that the mitochondrial genome pool can harbor signals that are not discernible in lower resolution data sets. In fact, the patterns would have not been revealed without dissection of the complete genome data into local (“Finn-characteristic”) and non-local components. The contemporary complete mtDNA genome data set showed diversity values, haplogroup distribution, and (deduced) past population sizes similar to those of many other European populations. However, segregating the sequences based on their inclusion in “local” and “non-local” haplogroups revealed diversity patterns in our model population of Finland. This approach provided an opportunity for closer assessment of the maternal past and yielded results that show closer overall resemblance to those obtained with other markers such as the Y-chromosome.



**Figure 1.** Median age estimates with 95% highest posterior density for Finn-characteristic subhaplogroups.



**Figure 2.** Effective population size ( $N_e$ ) comparison for Finn-characteristic (dark blue,  $N = 281$ ) and non-Finn-characteristic (orange,  $N = 562$ ) sequences. The larger graph zooms in on the last 11,000 years. In both graphs on the X-axis is time as years before present and on the Y-axis is effective population size represented on a logarithmic scale. The continuous center lines represent the mean for the  $N_e$ , and the dotted lines are the 95% highest posterior density intervals.

Interestingly, a large number of the haplogroups in the data is actually rare elsewhere in Europe. It is notable that these Finn-characteristic haplogroups, in which >75% of the mtDNAs were observed only in Finland, are at different depths in the global mtDNA phylogeny (ranging from e.g. V8 to H13a1a1d1). These lineages may have i) arisen in Finland and drifted to considerable frequency locally (autochthonous lineages); ii) arisen somewhere else but survived only in Finland by chance; and/or iii) these lineages are present in other populations, but have not been sampled. There is insufficient evidence to indicate which of these possibilities (or combinations thereof) best explain the haplogroups, and hence the term “Finn-characteristic” was applied instead of the term “autochthonous”

that would imply endemic history in Finland. However, the latter alternative (non-representative sampling) was considered unlikely given that the analysis was performed on the (sub)haplogroup level and that the number of complete mtDNA sequences recorded in GenBank is rather high (~30,000). When considering this bias it is worth noting that, as estimated from the Human Mitochondrial DataBase, Europeans are overrepresented in the GenBank (Supplementary Table S1), which should render the latter possibility less likely.

Another conspicuous difference is seen in the estimated past population sizes among the total data, Finnish, and non-Finnish haplogroups. The total data show development which is in concordance with previously published patterns for many European populations (including Finland<sup>50</sup>), or the Neolithic-associated H-haplogroup<sup>51</sup>. The deduced population sizes started to increase strongly in early Holocene c. 10,000–8,000 ybp, with an even sharper rise during the last few centuries. This increase has been associated with the Neolithic revolution and surplus nutrition produced after adopting agriculture. A general notion is that the hunter-gatherer populations show signatures of constant population sizes, different from the exponential growth patterns in the Neolithic cultures. A “Neolithic” pattern also was observed using the (Finnish) sequences in non-Finn-characteristic haplogroups. The Finn-characteristic haplogroups, however, showed a starkly different pattern, with rapidly decreasing population sizes throughout most of the Holocene, until a period of steep growth that started only as late as around 1,000 ybp. This difference between ‘local’ and ‘immigrant’ BSP curves is statistically significant. The results also suggest that a substantial reduction of population size dating took place around 2,000–1,000 years before present. This decrease in the estimated effective population sizes also was observed with the more stringent 90% cut-off limit used to delineate the Finnish haplogroups (Supplementary Fig. S5).

The population size reduction appears clear in the BSPs based on the Finn-characteristic haplotypes. Nevertheless, the possibility of this signal being an artefact must be considered. The BSP method assumes panmictic population, and violations, especially population structure, may produce similar signals as those of reductions in population size<sup>52–54</sup>. The magnitude of this problem obviously depends on the level of divergence between the (ignored) subpopulations, which in case of modern human mtDNA data usually is relatively small in Europe. However, a number of studies have demonstrated genetic substructure in Finland (see Introduction). In order to explore the possibility of spurious population change signals caused by a structure effect, the BSPs were constructed for modern HVR1 + 2 data sampled in the identified “subpopulations” in Finland<sup>9</sup>. The results (Supplementary Fig. S6) do not indicate a notable structure effect. The BSPs for the two subpopulations and the pooled population are similar, and none of them show population decline of the same magnitude that is observed in the Finn-characteristic haplotypes. This observation would tend to support that the overall declining pattern is real.

Two additional observations in the obtained graphs should be noted. First, the dip in the BSPs of Finn-characteristic haplotypes before the final population expansion must be treated with some caution, as similar patterns have been reported, e.g., from the Pacific herring (see Discussion in ref. 55). Secondly, it is worth noting that the values of estimated effective population may not be “real”, but can be treated as trends. The effective population size can be defined in a number of ways, and the issue is complex<sup>56</sup>.

Assuming the obtained results reflect the past population history, it is likely that both the high proportion of Finn-characteristic haplogroups and the reduced past population sizes arise from a small number of founders and relative genetic isolation for a rather long period of time. High levels of drift due to small population sizes and isolation-by-distance within Finland may have contributed strongly to the genetic diversity. While population bottlenecks have been suggested as a causal factor for Y-chromosomal and autosomal diversity patterns in Finland, and observations also in other types of data<sup>57</sup>, this pattern has rarely been deduced from mitochondrial DNA data. An exception is the study by Sajantila *et al.* 1996<sup>6</sup>, in which a bottleneck, dated to around 3,900 ybp, was inferred from the mitochondrial control region data. Sajantila *et al.* suggested that this bottleneck has influenced mtDNA diversity, but that mutations at rapidly evolving sites have generated diversity to levels observed in other parts of Europe. It is notable that, as in this study herein, the bottleneck was only visible after restricting the data into a subset of the total based on the variability of sites. A similar approach has been successfully employed by Fagundes *et al.* 2008<sup>2</sup>.

Although the approaches by Sajantila *et al.* 1996<sup>6</sup> and this study are in principle similar in observing patterns only in a restricted data set, they do convey different explanations. Considering the evolutionary forces, Sajantila *et al.* (1996) offer mutation as an explanation for the discrepancy between the high mitochondrial and lower Y-chromosomal diversity in Finland. However, there are several factors that may limit the role of mutation. First, the differences between Y-chromosomal and mtDNA mutation rates are not as large as the diversity differences between the two markers. Second, the use of pedigree rates has been shown to result in overestimates when considering data in a population sample<sup>58</sup>. Third, observing lower diversity in slowly evolving sites is likely to result from restricted haplogroup distribution, which is not a particularly conspicuous feature of the Finnish mtDNA pool. Alternatively, the current results, showing lower diversity (and hence population sizes) in the Finn-characteristic haplogroups, would suggest migration, instead of mutation, as an important source of relatively high mtDNA diversity in modern Finns. This notion would be in line with some simulation studies emphasizing constant, low-level gene flow in Finland<sup>59, 60</sup>.

The estimated haplogroup ages may shed light on the past events in Finland. Diversity within the non-Finnish haplogroups was shaped largely by distant environmental conditions and human population histories and can tell little about the Finns. Instead, the ages of the Finn-characteristic haplogroups range from 858 [57; 2,710] years for haplogroup W1b to 5,883 [2,708; 8,923] years for haplogroup U5b1b2\*. In general, there appears to be two loose and largely overlapping clusters among the Finn-characteristic haplogroups: the first between 1,000–2,000 ybp and the second around 3,300–5,500 ybp. The age of the older cluster coincides temporally with the arrival of the Corded-Ware culture and, notably, the spread of agriculture in Finland. The arrival and spread of agriculture, temporally corresponding with the age estimates for most of the haplogroups characteristic of Finns, might be a sign of population size increase enabled by the new mode of subsistence, resulting in reduced drift



and accumulation of genetic diversity in the population. Despite the documented demic diffusion associated with the Neolithic transition (see refs 61 and 62), it is likely that cultural diffusion also has taken place (see ref. 63). This pattern would explain the observed 'Neolithic' increase of diversity in both hunter-gatherer (U and V) and farmer (H, J and K) associated Finn-characteristic haplogroups. It should be noted that our data set might be slightly biased towards northern Finland, where the Corded-Ware culture did not reach. Note that the (Mesolithic/Neolithic) classification of haplogroup V is contentious, but here it is included in the hunter-gatherer pool due to its high frequency among the Saami, the indigenous reindeer herders in northern Finland. While the mtDNA diversity of the Saami is not the focus of the study herein, it is important to note that their history in Finland predates that of Finns<sup>64</sup> and that there has been (mtDNA) admixture between the Saami and the Finns (Evidence for mtDNA admixture between the Finns and the Saami in Meinilä *et al.*)<sup>65</sup>.

Another insight in the past population sizes in Finland is based on radiocarbon-dated archaeological findings in different time periods. These analyses suggest two prehistoric population peaks in Finland, the Stone Age peak (c. 5,500 ybp) and the Metal Age peak (~1,500 ybp)<sup>57, 66, 67</sup>. Both of these peaks were followed by a population decline, which appears to have reached its ebb around 3,500 ybp. These developments are not distinguishable in the BSPs. However, these ages correspond well to the two haplogroup age clusters described above. The presumably less severe Iron Age population bottleneck seen in the archaeological data, 1,500–1,300 ybp<sup>57, 67</sup>, temporally coincides with the population size reduction visible for the Finn-characteristic subhaplogroups.

## Conclusion

The palimpsest nature of today's genetic diversity may hide retrievable signals of past demography, only visible in subsets of data. In the case of Finland, dissection of the total variation into local and non-local variation components, enable visualization of demographic history patterns that are compatible with results deduced based on diverse data. These results also help to understand the apparent diversity differences between mitochondrial, Y-chromosomal and autosomal markers described in the Finnish population. Such methods that allow the deduction of the past from modern genetic data are especially helpful in regions where ancient DNA studies are limited due to paucity of archaeological bone material.

## References

1. Jobling, M. A. The impact of recent events on human genetic diversity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 793–799 (2012).
2. Fagundes, N. J., Kanitz, R. & Bonatto, S. L. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One* **3**, e3157 (2008).
3. Peltonen, L., Pekkarinen, P. & Aaltonen, J. Messages from an isolate: lessons from the Finnish gene pool. *BIOLOGICAL CHEMISTRY HOPPE SEYLER* **376**, 697–704 (1995).
4. Palo, J. U., Ulmanen, I., Lukka, M., Ellonen, P. & Sajantila, A. Genetic markers and population history: Finland revisited. *Eur. J. Hum. Genet.* **17**, 1336–1346 (2009).
5. Lao, O. *et al.* Correlation between genetic and geographic structure in Europe. *Current Biology* **18**, 1241–1248 (2008).
6. Sajantila, A. *et al.* Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl. Acad. Sci. USA* **93**, 12035–12039 (1996).
7. Lappalainen, T. *et al.* Regional differences among the finns: A Y-chromosomal perspective. *Gene* **376**, 207–215 (2006).
8. Salmela E. *et al.* Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in Northern Europe. *PLoS ONE* **3** (2008).
9. Neuvonen, A. M. *et al.* Vestiges of an Ancient Border in the Contemporary Genetic Diversity of North-Eastern Europe. *PLoS ONE* **10**, 1–19 (2015).
10. Sistonen, J. *et al.* Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. *Pharmacogenet Genomics* **19**, 170–179 (2009).
11. Neuvonen, A. M., Palo, J. U. & Sajantila, A. Post-mortem ABCB1 genotyping reveals an elevated toxicity for female digoxin users. *Int. J. Legal Med.* **125**, 265–269 (2011).
12. Torroni, A. *et al.* Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850 (1996).
13. Hedman, M. *et al.* Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Sci. Int.* **172**, 171–178 (2007).
14. Lappalainen, T. *et al.* Migration waves to the baltic sea region. *Ann. Hum. Genet.* **72**, 337–348 (2008).
15. Attimonelli, M. *et al.* HmtDB, a human mitochondrial genomic resource based on variability studies supporting population genetics and biomedical research. *BMC Bioinformatics* **6**(Suppl 4), S4 (2005).
16. Rubino, F. *et al.* HmtDB, a genomic resource for mitochondrion-based human variability studies. *Nucleic Acids Res.* **40**, D1150–9 (2012).
17. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
18. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147–147 (1999).
19. Finnilä, S., Lehtonen, M. S. & Majamaa, K. Phylogenetic network for European mtDNA. *The American Journal of Human Genetics* **68**, 1475–1484 (2001).
20. Soini, H. K., Moilanen, J. S., Vilmi-Kerala, T., Finnilä, S. & Majamaa, K. Mitochondrial DNA variant m.15218A>G in Finnish epilepsy patients who have maternal relatives with epilepsy, sensorineural hearing impairment or diabetes mellitus. *BMC Med. Genet.* **14**, 73–2350–14–73 (2013).
21. Soini, H. K., Moilanen, J. S., Finnilä, S. & Majamaa, K. Mitochondrial DNA sequence variation in Finnish patients with matrilineal diabetes mellitus. *BMC Res. Notes* **5**, 350-0500-5-350 (2012).
22. Raule, N. *et al.* The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. *Aging cell* **13**, 401–407 (2014).
23. Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics* **90**, 675–684 (2012).
24. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
25. Vianello, D. *et al.* HAPLOFIND: A New Method for High-Throughput mtDNA Haplogroup Assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
26. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* (2016).

27. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
28. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
29. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
30. Lanfear, R., Calcott, B., Ho, S. Y. & Guindon, S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
31. Soares, P. *et al.* The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2012).
32. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
33. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences* **17**, 57–86 (1986).
34. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
35. Kass, R. E., Carlin, B. P., Gelman, A. & Neal, R. M. Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician* **52**, 93–100 (1998).
36. Rambaut, A., Suchard, M., Xie, D. & Drummond, A. Tracer v1. 6. Computer program and documentation distributed by the author, website <http://beast.bio.ed.ac.uk/Tracer> [accessed 27 July 2014] (2014).
37. Newton, M. A. & Raftery, A. E. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*. 3–48 (1994).
38. Rieux, A. *et al.* Improved calibration of the human mitochondrial clock using ancient genomes. *Mol. Biol. Evol.* **31**, 2780–2792 (2014).
39. Molak, M., Lorenzen, E. D., Shapiro, B. & Ho, S. Y. Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Mol. Biol. Evol.* **30**, 253–262 (2013).
40. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
41. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
42. Tang, H., Siegmund, D. O., Shen, P., Oefner, P. J. & Feldman, M. W. Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics* **161**, 447–459 (2002).
43. Mishmar, D. *et al.* Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* **100**, 171–176 (2003).
44. Soares, P. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics* **84**, 740–759 (2009).
45. Ho, S. Y. & Endicott, P. The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am. J. Hum. Genet.* **83**, 142 (2008).
46. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology* **23**, 553–559 (2013).
47. Korkko, J., Annunen, S., Pihlajamaa, T., Prockop, D. J. & Ala-Kokko, L. Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. *Proc. Natl. Acad. Sci. USA* **95**, 1681–1685 (1998).
48. Pinhasi, R., Thomas, M. G., Hofreiter, M., Currat, M. & Burger, J. The genetic history of Europeans. *Trends Genet.* **28**, 496–505 (2012).
49. Samuels, D. C., Boys, R. J., Henderson, D. A. & Chinnery, P. F. A compositional segmentation of the human mitochondrial genome is related to heterogeneities in the guanine mutation rate. *Nucleic Acids Res.* **31**, 6043–6052 (2003).
50. Zheng, H., Yan, S., Qin, Z. & Jin, L. MtDNA analysis of global populations support that major population expansions began before Neolithic Time. *Scientific reports* **2** (2012).
51. Fu, Q., Rudan, P., Pääbo, S. & Krause, J. Complete mitochondrial genomes reveal Neolithic expansion into Europe. *PLoS One* **7**, e32473 (2012).
52. Chikhi, L., Sousa, V. C., Luisi, P., Goossens, B. & Beaumont, M. A. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186**, 983–995 (2010).
53. Mazet, O., Rodriguez, W., Grusea, S., Boitard, S. & Chikhi, L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference. *Heredity* **116**, 362–371 (2016).
54. Heller, R., Chikhi, L. & Siegmund, H. R. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One* **8**, e62992 (2013).
55. Grant, W. S., Liu, M., Gao, T. & Yanagimoto, T. Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). *Mol. Phylogenet. Evol.* **65**, 203–212 (2012).
56. Templeton, A. & Read, B. In *Conservation genetics* 91–105 (Springer, 1994).
57. Tallavaara, M., Pesonen, P. & Oinonen, M. Prehistoric population history in eastern Fennoscandia. *Journal of Archaeological Science* **37**, 251–260 (2010).
58. Howell, N. *et al.* The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *The American Journal of Human Genetics* **72**, 659–670 (2003).
59. Sundell, T., Heger, M., Kammonen, J. & Onkamo, P. Modelling a Neolithic population bottleneck in Finland: a genetic simulation. *Fennoscandia Archaeologica* **27**, 3–19 (2010).
60. Sundell, T., Kammonen, J., Heger, M., Palo, J. U. & Onkamo, P. *Retracing prehistoric population events in Finland using simulation* (Archaeology in the digital era: papers from the 40th annual conference of computer applications and quantitative methods in archaeology (CAA), Southampton, 26–29 March 2012, Amsterdam University Press, 2014).
61. Chikhi, L., Nichols, R. A., Barbujani, G. & Beaumont, M. A. Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci. USA* **99**, 11008–11013 (2002).
62. Balaesque, P. *et al.* A predominantly neolithic origin for European paternal lineages. *PLoS Biol* **8**, e1000285 (2010).
63. Fort, J. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proc. Natl. Acad. Sci. USA* **109**, 18669–18673 (2012).
64. Ingman, M. & Gyllensten, U. A recent genetic link between Sami and the Volga-Ural region of Russia. *European Journal of Human Genetics* **15**, 115–120 (2007).
65. Meinila, M., Finnila, S. & Majamaa, K. Evidence for mtDNA admixture between the Finns and the Saami. *Hum. Hered.* **52**, 160–170 (2001).
66. Tallavaara, M. & Seppä, H. Did the mid-Holocene environmental changes cause the boom and bust of hunter-gatherer population size in eastern Fennoscandia? *The Holocene*, 0959683611414937 (2011).
67. Oinonen, M., Pesonen, P. & Tallavaara, M. Archaeological radiocarbon dates for studying the population history in eastern Fennoscandia. *Radiocarbon* **52**, 393 (2010).
68. Pike, D. A., Barton, T. J., Bauer, S. L. & Kipp, E. B. mtDNA Haplogroup T Phylogeny Based on Full Mitochondrial Sequences. *J Genet Geneal* **6**, 1–24 (2010).

69. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
70. Stoljarova, M., King, J. L., Takahashi, M., Aaspõllu, A. & Budowle, B. Whole mitochondrial genome genetic diversity in an Estonian population sample. *Int. J. Legal Med.* **130**, 67–71 (2016).

## Acknowledgements

This work was funded by Finnish Cultural Foundation (<https://skr.fi/en>; SÖ), the Academy of Finland grant no. 133056 (<http://www.aka.fi/fi/A/>; SÖ, PO) and the Finnish Foundations' Professor Pool Grant (<http://www.professoripooli.fi>; Paulo Foundation; AS). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author Contributions

S.Ö., J.U.P. and P.O. conceived and designed the experiment; J.U.P., P.O. and A.S. supervised this study; S.Ö. and J.U.P. performed data collection and data analysis; P.O. helped with the data analysis; M.S. and B.B. provided the data; J.U.P. and S.Ö. wrote the manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-05673-7](https://doi.org/10.1038/s41598-017-05673-7)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017